An entropic inequality in finite Abelian groups analogous to the unified Brascamp-Lieb and Entropy Power Inequality

Chin Wa (Ken) Lau and Chandra Nair
Dept. of Information Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong (China)
Email:{kenlau,chandra}@ie.cuhk.edu.hk

Abstract

The doubling-followed-by-rotation trick to prove the extremality of Gaussian distributions has been a valuable tool in information theory. In particular, the above trick has been used to establish the Gaussian extremality of a family of inequalities that unifies the Entropy Power Inequality and the Brascamp-Lieb inequalities. Here, we develop a technique (similar to the one in the continuous case) to prove[1] the extremality of Haar distributions for a similar family of inequalities in finite Abelian groups.

I. Introduction

A. Background

Entropy Power Inequality (EPI) is a powerful tool that has found widespread applications in network information theory. It has been widely used to show the capacity region (for instance [Ber73], [WSS06]) in several multiuser information theory settings. Furthermore, various versions of this inequality have been formulated for discrete random variables. Shamai and Wyner, [SW90], established a discrete analog of EPI for the binary random variables. Harremoës and Vignaet, [HV03], discovered a discrete analog of EPI for a particular family of binomial random variables. Sharma, Das, and Muthukrishnan, [SDM11] based on the work of [HV03], establish another version of the discrete EPI. On the other hand, there have been several attempts to generalize Mrs. Gerber's Lemma (Wyner and Ziv [WZ73]); for example, Jog and Anantharam have shown a generalization of Mrs. Gerber's Lemma for random variables on the Abelian group with order 2^n [JA14].

Theorem 1 (Entropy Power Inequality [Sha48], [Sta59])**.** *Suppose X and Y are independent $\mathbb{R}^n$-valued random variables. The entropy power of X is defined as*

$$\mathcal{N}(X) = \frac{1}{2\pi e} e^{2h(X)/n},$$

where $h(X)$ is the differential entropy of X.

The Entropy Power Inequality states that

$$\mathcal{N}(X) + \mathcal{N}(Y) \leq \mathcal{N}(X+Y),$$

where the equality holds if and only if X and Y are Gaussians with proportional covariance matrices.

An equivalent dimension-independent form of the Entropy Power Inequality was formulated by Lieb [Lie78].

Theorem 2. *Suppose X and Y are independent $\mathbb{R}^n$-valued random variables. For any $\lambda \in [0,1]$, we have*

$$\inf_{X,Y:X\perp Y} h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y) \geq 0,$$

where the equality holds if and only if X and Y are Gaussians with proportional covariance matrices.

In other words, the functional

$$f(\mu_X, \mu_Y) : h(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda h(X) - (1-\lambda)h(Y),$$

where $X \sim \mu_X$ and $Y \sim \mu_Y$ are independent random variables and are minimized by Gaussians with proportional covariance matrices.

Similarly, the Brascamp-Lieb inequality (BLI) [BL76] is a family of functional inequalities at the intersection of information and functional inequalities. Special cases of the BLI include Hölder's inequality, the Loomis-Whitney inequality, the Prékopa-Leindler inequality, and sharp forms of Young's convolution inequalities [BCCT08]. One of the central results here is that the

[1] A conference version of the results in the first half appeared in the International Symposium on Information Theory, 2024 [LN24].

optimal constants can be computed by restricting to Gaussian distributions. Recently, in [AJN22], the following theorem was proved that unified the family of Brascamp-Lieb inequalities and the Entropy-Power inequality.

Theorem 3 (Unified EPI and BLI, [AJN22])**.** *Let* $(\mathbf{A}, \mathbf{c}, \mathbf{r}, \mathbf{d})$ *be a BL-EPI datum. Define*

$$M_g := \sup_{Z \in \mathcal{P}_g(\mathbf{r})} \sum_{i=1}^{k} d_i h(Z_i) - \sum_{j=1}^{m} c_j h(A_j Z).$$

Then for any $X \in \mathcal{P}(\mathbf{r})$*, the following inequality holds:*

$$\sum_{i=1}^{k} d_i h(X_i) - \sum_{j=1}^{m} c_j h(A_j X) \leq M_g.$$

Remark 1. The readers are encouraged to look at [AJN22] for a precise definition of BL-EPI datum and $\mathcal{P}(\mathbf{r})$. The main point is that $\mathcal{P}_g(\mathbf{r})$ restricts the distributions in $\mathcal{P}(\mathbf{r})$ to Gaussian distributions, and Z's are Gaussian random variables. For this paper, it suffices to note that $\{d_i\}$ and $\{c_j\}$ are positive constants, and $X_1, X_2, \ldots, X_k$ are mutually independent random vectors. It is also worth noting that the same proof (of Gaussian extremality) goes through if one imposes covariance constraints, $\mathbb{E}[X_i X_i^T] \preceq K_i$, on the independent random vectors.

The above inequality was proved using the doubling and rotation idea (as developed in [GN14]), a technique that differs from the previous proof methods of the Entropy Power Inequality. Our main result (Theorem 6) is a discrete analog (in finite Abelian groups) of Theorem 3. Further, we demonstrate that the proof technique in [AJN22] can be essentially mimicked (modulo some differences in the technical arguments) in this setting. The proof technique in [AJN22] can be summarized (pushing some technical conditions under the carpet) as follows: Gaussian optimality was deduced by using the sub-additivity of an entropic functional and by using this to show that rotated forms of the optimizers are independent. This implied that the optimizers must be Gaussian by applying the Darmois-Skitovich theorem.

Theorem 4 (Darmois-Skitovich theorem [Dar53], [Ski53])**.** *Let* $X_1, \ldots, X_n$ *be independent random variables. Let* $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ *be non-zero constants for each coordinate. If the linear statistics* $L_1 = \sum_{i=1}^{n} \alpha_i X_i$ *and* $L_2 = \sum_{i=1}^{n} \beta_i X_i$ *are independent, then all random variables* $X_1, \ldots, X_n$ *are Gaussians.*

A finite Abelian group analog of this was discovered by Feldman [Fel99].

Theorem 5 (Feldman [Fel99])**.** *Let* $\mathbb{G}$ *be a finite Abelian group, and* X_1, X_2 *be independent random variables with values in* $\mathbb{G}$*. Let* $\alpha_1, \alpha_2, \beta_1, \beta_2$ *be automorphisms of the group* $\mathbb{G}$*. Then if the linear statistics* $L_1 = \alpha_1(X_1) + \alpha_2(X_2)$ *and* $L_2 = \beta_1(X_1) + \beta_2(X_2)$ *are independent, then* X_1 *and* X_2 *are shifts of a Haar distribution of some subgroup* $\mathbb{H}$ *of* $\mathbb{G}$*, or equivalently,* X_1 *and* X_2 *are uniform distributions on a coset of some subgroup* $\mathbb{H}$ *of the group* $\mathbb{G}$*.*

Remark 2. The uniform distribution on a coset of some subgroup $\mathbb{H}$ of a finite Abelian group $\mathbb{G}$ has very similar properties to that of Gaussians in the respective from the above theorem. By shifting the mean, we see that Haar distributions (uniform distributions) on subgroups play an analogous role to Gaussian distributions.

Therefore, it is natural to guess that Gaussians can be replaced by uniform distributions on a coset (corresponding to a shift in the mean) of some subgroup (or shifts of Haar distributions) when working in finite Abelian groups. However, while this intuition is correct, we show a way to overcome some technical hassles (different from the continuous case) in our proof. Furthermore, just like the rotation trick in the continuous case, we believe this argument can find several other applications to establish the optimality of Haar distributions.

Notation: We use $(\mathbb{G}, +)$ or $\mathbb{G}$ to denote a finite Abelian group. We use $|A|$ to denote the cardinality of a finite set A and $\mathrm{support}(p_X)$ to denote the support of p_X.

II. Main

Theorem 6. *Let* $X_1, \ldots, X_n$ *be independent random variables taking values in some subgroup* $\mathbb{H}_1, \ldots, \mathbb{H}_n$ *of a finite Abelian group* $\mathbb{G}$*. Let* $a_1, \ldots, a_n$*, and* $b_1, \ldots, b_\ell$ *be positive constants, and* $c_{i,j}^{(1)}, \ldots, c_{i,j}^{(m_j)}$ *be integers. Then, the following optimization problem*

$$\max_{\prod_{i=1}^{n} p_{X_i}} \sum_{i=1}^{n} a_i H(X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i \right),$$

has an optimizer $(X_1^*, \ldots, X_n^*)$ *of the form, each* X_i^* *has an uniform distribution on a coset of a subgroup* $\mathbb{K}_i \subseteq \mathbb{H}_i$*.*

Remark 3. The following points are worth noting:

1) One can relax the assumption on the sign of a_i. Note that, if any $a_k \leq 0$, it is immediate that an optimal choice is to set the corresponding X_k to be a constant random variable. To see this one observes that $H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i, \right) \geq H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | X_k\right)$.
2) Unlike the continuous case, where Lieb's formulation of EPI was known, the extremality of the uniform distribution of a coset of some subgroup for $a_1 H(X_1) + a_2 H(X_2) - H(X_1 + X_2)$ was not known. There have been conjectures (and some results), [JA14], of a similar flavor.

We establish the following lemma before providing proof of Theorem 6. This is the analogous result of the Darmois-Skitovich theorem we need in our proof.

Lemma 1. *Let X_A and X_B be two independent random variables taking values in some finite Abelian group $\mathbb{H}$. Let S denote the support of the probability distribution of X_B. Let $\mathbb{D}$ denote the subgroup generated by the pairwise differences of the elements of* $\mathrm{support}(X_B)$. *For $X_A + X_B$ to be independent of X_B, it is necessary and sufficient that $\mathrm{P}(X_A = h_1) = \mathrm{P}(X_A = h_2)$ whenever h_1, h_2 belong to the same coset of $\mathbb{D}$ (in other words, p_{X_A} is uniformly distributed conditioned on it taking values in a given coset of $\mathbb{D}$). Consequently $|\mathrm{support}(X_A)| = k|\mathbb{D}| \geq k|\mathrm{support}(X_B)|$ for some $k \in \mathbb{N}$ satisfying $1 \leq k \leq \frac{|\mathbb{H}|}{|\mathbb{D}|}$, and $k = 1$ only if X_A is uniformly distributed on a coset of $\mathbb{D}$.*

Proof. First, assume that X_A is uniform on the cosets of $\mathbb{D}$. Let T be a set of coset representatives, i.e., a transversal of the collection of cosets of $\mathbb{D}$. Therefore, any element $h \in H$ can be uniquely represented as $h = t + d$, for some $t \in T$ and $d \in \mathbb{D}$. If X_A is uniform on the cosets of $\mathbb{D}$, then $\mathrm{P}(X_A = h) = \mathrm{P}(X_A = t + d) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)$ for some arbitrary distribution on the transversal. If X_A and X_B are independent, note that $\mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\mathrm{P}(X_B = b)$.

On the other hand $\mathrm{P}(X_A + X_B = h + b) = \sum_{\hat{b} \in S} \mathrm{P}(X_A = h + b - \hat{b})\mathrm{P}(X_B = \hat{b})$. Since $b - \hat{b} \in D$, $h + b - \hat{b}$ belongs to the same coset as h. Therefore, for all $\hat{b}$, we have $\mathrm{P}(X_A = h + b - \hat{b}) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)$. Consequently, $\mathrm{P}(X_A + X_B = h + b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\sum_{\hat{b} \in S} \mathrm{P}(X_B = \hat{b}) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)$. Therefore $\mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b) = \frac{1}{|\mathbb{D}|}\mathrm{P}(T = t)\mathrm{P}(X_B = b) = \mathrm{P}(X_A + X_B = h + b)\mathrm{P}(X_B = b)$. This implies that $X_A + X_B$ is also independent of X_B.

Conversely, let us assume that X_A and X_B are independent, and additionally, $X_A + X_B$ is also independent of X_B. Therefore $\mathrm{P}(X_A + X_B = h + b)\mathrm{P}(X_B = b) = \mathrm{P}(X_A + X_B = h + b, X_B = b) = \mathrm{P}(X_A = h)\mathrm{P}(X_B = b)$. This implies that for all $b \in S$, we have $\mathrm{P}(X_A = h) = \mathrm{P}(X_A + X_B = h + b) = \sum_{\hat{b} \in S} \mathrm{P}(X_A = h + b - \hat{b})\mathrm{P}(X_B = \hat{b})$. Rewriting h as $h - b$, we see that $\mathrm{P}(X_A = h - b) = \sum_{\hat{b} \in S} \mathrm{P}(X_A = h - \hat{b})\mathrm{P}(X_B = \hat{b})$. Since the right-hand-side does not depend on b, we obtain that $\mathrm{P}(X_A = h - b_1) = \mathrm{P}(X_A = h - b_2)$, for all $b_1, b_2 \in S$ and $h \in \mathbb{H}$. Replacing $h - b_1$ by h, we note that $\mathrm{P}(X_A = h) = \mathrm{P}(X_A = h + b_1 - b_2)$. Since the pairwise differences $b_i - b_j$ generate $\mathbb{D}$, and from above p_{X_A} is invariant under a shift by a pairwise difference, it follows that p_{X_A} is invariant under a shift by an element in $\mathbb{D}$. In other words, X_A is uniform on the cosets of $\mathbb{D}$.

Finally note that $|\mathrm{support}(X_A)| = |\mathrm{support}(T)||\mathbb{D}|$, and $|\mathrm{support}(X_A)| = |\mathbb{D}|$ only if T is a constant random variable, implying that X_A is uniform on a coset of $\mathbb{D}$. We also have that $|\mathbb{D}| \geq |\mathrm{support}(X_B)|$, since $b \mapsto b - b_0$ is an injection from $\mathrm{support}(X_B)$ to $\mathbb{D}$, where b_0 is an arbitrary fixed element from $\mathrm{support}(X_B)$. □

Remark 4. The proof is similar to that in [Tao10, Section 5]. In [Tao10], X_A and X_B are assumed to be identically distributed.

A. Proof of Theorem 6

The first step in proving the optimality of the uniform distribution of a coset of some subgroup is to identify a sub-additive functional. To this end, given an n-tuple of distributions $(p_{X_1}, \ldots, p_{X_n})$, such that X_i has support on $\mathbb{H}_i$, let us define:

$$F(X_1, \ldots, X_n) := \sup_{\substack{p_{U|X_1,\ldots,X_n}: \\ p_{X_1,\ldots,X_n|U} = \prod_{i=1}^n p_{X_i|U}}} \sum_{i=1}^n a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i | U\right).$$

Observe that the maximum value of $F(X_1, \ldots, X_n)$ is the same as the value of the optimization problem in Theorem 6, as the average is always less than the maximum (the other direction is immediate by taking $X_1, \ldots, X_n$ to be mutually independent and U to be a constant).

Remark 5. This is essentially the same function as the one employed in [AJN22].

Now consider an n-tuple of distributions $(p_{X_1,\hat{X}_1}, \ldots, p_{X_n,\hat{X}_n})$, such that $(X_i, \hat{X}_i)$ has support on $\mathbb{H}_i \times \hat{\mathbb{H}}_i$, let us define (ignoring the abuse of notation):

$$F((X_1, \hat{X}_1), \ldots, (X_n, \hat{X}_n)) := \sup_{\substack{p_{U|(X_1,\hat{X}_1),\ldots,(X_n,\hat{X}_n)}: \\ p_{(X_1,\hat{X}_1),\ldots,(X_n,\hat{X}_n)|U} = \prod_{i=1}^n p_{(X_i,\hat{X}_i)|U}}} \sum_{i=1}^n a_i H(X_i, \hat{X}_i|U)$$

$$- \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i, \sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i | U\right).$$

Observe that

$$\begin{aligned}
&\sum_{i=1}^{n} a_i H(X_i, \hat{X}_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i, \sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i | U\right) \\
&\quad = \sum_{i=1}^{n} a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i | U\right) \\
&\quad + \sum_{i=1}^{n} a_i H(\hat{X}_i|U, X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i | U, \sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i\right) \\
&\quad \overset{(a)}{=} \sum_{i=1}^{n} a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i | U\right) \\
&\quad + \sum_{i=1}^{n} a_i H(\hat{X}_i|U, \mathbf{X}) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i | U, \sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i\right) \\
&\quad \overset{(b)}{\leq} \sum_{i=1}^{n} a_i H(X_i|U) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i | U\right) \\
&\quad + \sum_{i=1}^{n} a_i H(\hat{X}_i|U, \mathbf{X}) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i | U, \mathbf{X}\right) \\
&\quad \overset{(c)}{\leq} F(X_1, \ldots, X_n) + F(\hat{X}_1, \ldots, \hat{X}_n).
\end{aligned}$$

In the above $\mathbf{X} = (X_1, \ldots, X_n)$. Equality (a) follows, as conditioned on U, $\{(X_i, \hat{X}_i)\}$ are mutually independent and equality (b) follows from data-processing inequality as $(U, \sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i) \to (U, \mathbf{X}) \to (U, \sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i)$ is Markov. Finally inequality (c) follows since conditioned on U, the random variables $\{X_i\}$ are mutually independent, and conditioned on $(U, \mathbf{X})$, the random variables $\{\hat{X}_i\}$ are mutually independent.

Remark 6. The next step in the proof (in the continuous case) is to argue that rotated versions of two independent copies of the maximizers are independent. In the continuous case, this involves showing the existence of the maximizers and then (sometimes) considering a perturbed function to deduce the independence of the rotated versions. In the finite alphabet case, the existence of the maximizers is immediate but one still needs to consider a perturbed function to deduce the independence.

In the next part of the proof, we will argue that certain linear forms of the maximizer are independent. To this end, consider the two maximization problems listed below:

$$\max_{\prod_{i=1}^{n} p_{X_i}} \sum_{i=1}^{n} a_i H(X_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i\right),$$

$$\max_{\prod_{i=1}^{n} p_{\hat{X}_i}} \sum_{i=1}^{n} a_i H(\hat{X}_i) - \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} \hat{X}_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} \hat{X}_i\right) - \sum_{i=1}^{n} \epsilon H(\hat{X}_i).$$

In the above two problems, the random variables X_i and $\hat{X}_i$ are assumed to take values in the subgroup $\mathbb{H}_i$. Let $(X_1^*, \ldots, X_n^*)$ and $(\hat{X}_{1,\epsilon}^*, \ldots, \hat{X}_{n,\epsilon}^*)$ be maximizers of the two optimization problems respectively and V, V_ϵ be the maximum value attained by the two optimization problems. Further, let us assume that among all possible maximizers of the first problem, $(X_1^*, \ldots, X_n^*)$ minimizes the function $\prod_{i=1}^{n}(1 + |\text{support}(X_i)|)$.

It is immediate that $V_\epsilon \to V$ and $\epsilon \to 0$ (as the difference between the objective functions at any point is bounded by $\epsilon \left(\sum_{i=1}^{n} \log |\mathbb{H}_i|\right)$. Furthermore, by the compactness of the probability simplex and continuity of the function, we know that there is a sequence of maximizers $(\hat{X}_{1,\epsilon_m}^*, \ldots, \hat{X}_{n,\epsilon_m}^*)$ that converge to a maximizer of the first optimization problem.

Finally, we define

$$\begin{aligned}
F_\epsilon(X_1, \ldots, X_n) := \sup_{\substack{p_{U|X_1,\ldots,X_n}: \\ p_{X_1,\ldots,X_n|U} = \prod_{i=1}^{n} p_{X_i|U}}} &\sum_{i=1}^{n} a_i H(X_i|U) \\
&- \sum_{j=1}^{\ell} b_j H\left(\sum_{i=1}^{n} c_{i,j}^{(1)} X_i, \ldots, \sum_{i=1}^{n} c_{i,j}^{(m_j)} X_i | U\right) - \sum_{i=1}^{n} \epsilon H(X_i|U).
\end{aligned}$$

We have $F_\epsilon(X_1,\ldots,X_n) \le V_\epsilon$.

Observe that by taking independent copies of the maximizers $(X_1^*,\ldots,X_n^*)$ and $(\hat{X}_{1,\epsilon}^*,\ldots,\hat{X}_{n,\epsilon}^*)$, we obtain

$$
\begin{aligned}
V + V_\epsilon &= \sum_{i=1}^n a_i H(X_i^*) - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i^*\right) \\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_{i,\epsilon}^*) - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) - \sum_{i=1}^n \epsilon H(\hat{X}_{i,\epsilon}^*) \\
&\overset{(a)}{=} \sum_{i=1}^n a_i H(X_i^*, \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^n \epsilon H(\hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} X_i^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} X_i^*, \sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) \\
&\overset{(b)}{=} \sum_{i=1}^n a_i H(X_i^* + \hat{X}_{i,\epsilon}^*, \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^n \epsilon H(\hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*), \sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^*\right) \\
&= \sum_{i=1}^n a_i H(X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_{i,\epsilon}^* | X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^n \epsilon H(\hat{X}_{i,\epsilon}^* | X_i^* + \hat{X}_{i,\epsilon}^*) - \sum_{i=1}^n \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^* | \sum_{i=1}^n c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\overset{(c)}{\le} \sum_{i=1}^n a_i H(X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} (X_i^* + \hat{X}_{i,\epsilon}^*), \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} (X_i^* + \hat{X}_{i,\epsilon}^*)\right) \\
&\quad + \sum_{i=1}^n a_i H(\hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*) - \sum_{i=1}^n \epsilon H(\hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*) - \sum_{i=1}^n \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\quad - \sum_{j=1}^\ell b_j H\left(\sum_{i=1}^n c_{i,j}^{(1)} \hat{X}_{i,\epsilon}^*, \ldots, \sum_{i=1}^n c_{i,j}^{(m_j)} \hat{X}_{i,\epsilon}^* | \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*\right) \\
&\overset{(d)}{\le} F(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*) + F_\epsilon(\hat{X}_{1,\epsilon}^*, \ldots, \hat{X}_{n,\epsilon}^*) - \sum_{i=1}^n \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) \\
&\overset{(e)}{\le} V + V_\epsilon - \sum_{i=1}^n \epsilon I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*).
\end{aligned}
$$

Here $\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*$ stands for the vector $(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*)$. In the above, equality (a) follows from the independence of $\mathbf{X}^*$ and $\hat{\mathbf{X}}_\epsilon^*$ and equality (b) follows from $H(X_1, X_2) = H(X_1 + X_2, X_2)$. Equality (c) follows from data-processing and the independence of the components of $(\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*)$, and (d) follows from the definition of F and F_ϵ as elaborated next. Note that $(X_1^* + \hat{X}_{1,\epsilon}^*, \ldots, X_n^* + \hat{X}_{n,\epsilon}^*)$ satisfies the support constraints and is a valid input for the function F (with U taken to be a constant). Now take $U = \mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*$ and use independence of the components of $(\mathbf{X}^* + \hat{\mathbf{X}}_\epsilon^*)$ to justify that this choice is a valid extension$p_{U|\hat{\mathbf{X}}_\epsilon^*}$ in the definition of F_ϵ. Finally, we note that the maximum of F and F_ϵ are V and V_ϵ to justify the inequality (e).

For $\epsilon > 0$, note that the above manipulations imply that $I(\hat{X}_{i,\epsilon}^*; X_i^* + \hat{X}_{i,\epsilon}^*) = 0$ using the non-negativity of mutual information, or in other words, that $X_i^* + \hat{X}_{i,\epsilon}^*$ is independent of $\hat{X}_{i,\epsilon}^*$. Since X_i^* was independent of $\hat{X}_{i,\epsilon}^*$ by construction, note that we can apply Lemma 1 to deduce that the distribution of X_i^* is uniform on the cosets of $\mathbb{D}_{i,\epsilon}$. Here $\mathbb{D}_{i,\epsilon}$ is the subgroup of $\mathbb{H}_i$ generated by the pairwise differences of the support of $\hat{X}_{i,\epsilon}^*$. Further $|\mathrm{support}(X_i^*)| = k_{i,\epsilon}|\mathbb{D}_{i,\epsilon}|$ for some $k_{i,\epsilon} \in \mathbb{N}$ satisfying $1 \le k_{i,\epsilon} \le \frac{|\mathbb{H}_i|}{|\mathbb{D}_{i,\epsilon}|}$.

As argued earlier, we have a sequence of optimizers $\hat{\mathbf{X}}^*_{\epsilon_m}$ such that as $\epsilon_m \downarrow 0$ and $\hat{\mathbf{X}}^*_{\epsilon_m}$ converges to a maximizer, say $\tilde{\mathbf{X}}^*$, of the problem with $\epsilon = 0$. Now, we have for any $\epsilon > 0$,

$$\prod_{i=1}^{n}(1 + k_{i,\epsilon}|\mathbb{D}_{i,\epsilon}|) = \prod_{i=1}^{n}(1 + |\text{support}(X_i^*)|) \leq \prod_{i=1}^{n}(1 + |\text{support}(\tilde{X}_i^*)|)$$
$$= \lim_{m\to\infty}\prod_{i=1}^{n}(1 + |\text{support}(\hat{X}^*_{i,\epsilon_m})|) \leq \lim_{m\to\infty}\prod_{i=1}^{n}(1 + |\mathbb{D}_{i,\epsilon_m}|).$$

The second assertion holds because we assumed that $\mathbf{X}^*$ minimizes $\prod_{i=1}^{n}(1 + |\text{support}(X_i)|)$ among all the maximizers of the optimization problem. This forces, for each $1 \leq i \leq n$, the sequence $k_{i,\epsilon_m} \to 1$ as $m \to \infty$. Therefore for some large, enough m, have $k_{i,\epsilon_m} = 1$ for all i, where $1 \leq i \leq m$. Therefore, again invoking Lemma 1, we see that X_i^* is uniformly distributed on some coset of a subgroup $\mathbb{D}_{i,\epsilon_m} \subseteq \mathbb{H}_i$. This completes the proof of Theorem 6.

Remark 7. Note that the above argument also establishes some properties of the maximizers of the optimization problem. Suppose $\mathbf{X}_a^*$ is another maximizer such that $\prod_{i=1}^{n}(1+|\text{support}(X^*_{a,i})|) > \prod_{i=1}^{n}(1+|\text{support}(X_i^*)|)$. Then, the above argument implies that one cannot have a sequence of maximizers of the perturbed problem that converges to $\mathbf{X}_a^*$.

III. A RELATED ARGUMENT FOR THE OPTIMALITY OF UNIFORM DISTRIBUTION

Optimality of uniform distribution in a discrete setting for an information functional has recently been established by Gowers, Green, Manners, and Tao [GGMT23].

A. Basic Definitions

Definition 1 (Independent Entropic Ruzsa distance, [GGMT23])**.** Suppose X, Y are $\mathbb{G}$-valued random variables. The independent entropic Ruzsa distance between X and Y is defined as

$$d(X, Y) = H(X' + Y') - \frac{1}{2}H(X) - \frac{1}{2}H(Y),$$

where X' and Y' are independent copies of X, Y.

Remark 8. This is sometimes defined as $H(X'-Y')-\frac{1}{2}H(X)-\frac{1}{2}H(Y)$. For groups with characteristic 2, these two definitions are equivalent. There is also another "entropic Ruzsa distance" (defined in [KLN23]) where

$$d(X, Y) = \max_{\Pi(p_x,p_y)} H(X' - Y') - \frac{1}{2}H(X) - \frac{1}{2}H(Y),$$

where $\Pi(p_x, p_y)$ denotes the set of couplings with fixed marginals. Note that none of the above definitions is a distance. When $p_x = p_y$, it does not hold that $d(X, Y) = 0$.

Lemma 2. *The independent entropic Ruzsa distance satisfies the triangle inequality, i.e.* $d(X, Z) \leq d(X, Y) + d(Y, Z)$.

Proof. Let (X, Y, Z) be independent. What we need to show is equivalent to

$$H(X + Z) + H(Y) \leq H(X + Y) + H(Y + Z).$$

This can be rewritten as

$$I(X; X + Y + Z) \leq I(X; X + Y) + I(Y; X + Y + Z).$$

By data-processing inequality, as $X \to X + Y \to X + Y + Z$ is Markov, $I(X; X + Y + Z) \leq I(X; X + Y)$ and the lemma follows. □

Definition 2 (Conditionally-Independent Entropic Ruzsa distance)**.** Suppose X, Y are $\mathbb{G}$-valued random variables. The conditionally-independent entropic Ruzsa distance between X and Y is defined as

$$d(X, Y|U) = H(X' + Y'|U) - \frac{1}{2}H(X|U) - \frac{1}{2}H(Y|U)$$

where $(U, X') \sim (U, X)$, $(U, Y') \sim (U, Y)$, and $X' \to U \to Y'$ is Markov.

Definition 3 (Polynomial Freiman-Ruzsa Functional)**.** [GGMT23, Equation 2.1] For any random variables X^0, Y^0 with support contained inside $\mathbb{G}$, a finite Abelian group with characteristic 2, define the functional

$$\tau(X, Y; X^0, Y^0) := \left(H(X + Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y)\right) + \eta\left(H(X + X^0) - \frac{1}{2}H(X) - \frac{1}{2}H(X^0)\right)$$
$$+ \eta\left(H(Y + Y^0) - \frac{1}{2}H(Y) - \frac{1}{2}H(Y^0)\right),$$

where X, Y, X^0, Y^0 are mutually independent. Here X, Y also take values in $\mathbb{G}$.

It was shown in [GGMT23, Proposition 2.1] that all minimizers of $\tau(X, Y)$ must be uniform distributions on a coset of a subgroup for all X^0, Y^0 with support in $\mathbb{G}$, when $\eta \leq \frac{1}{9}$.

A natural question to ask is whether there is a related super-additive function and whether one can use the machinery developed in the first part of the paper to deduce the optimality of the uniform distribution. The answer to the former part is yes, while the latter part seems to be not as straightforward.

Let us consider a slight modification of the above functional.

Definition 4 (Conditional PFR Functional)**.** Let X^0 and Y^0 be fixed $\mathbb{G}$-valued random variables. Suppose U, X, Y are $\mathbb{G}$-valued random variables. We require the triple $(U, X, Y), X^0, Y^0$ are independent. We define the conditional PFR functional as below

$$\begin{aligned}\tau(X, Y; X^0, Y^0|U) &= d(X, Y|U) + \eta d(X, X^0|U) + \eta d(Y, Y^0|U)\\ &= H(X' + Y'|U) - \frac{1+\eta}{2}H(X|U) - \frac{1+\eta}{2}H(Y|U)\\ &\quad + \eta H(X + X^0|U) + \eta H(Y + Y^0|U) - \frac{\eta}{2}H(X^0) - \frac{\eta}{2}H(Y^0)\end{aligned}$$

where $(U, X') \sim (U, X)$, $(U, Y') \sim (U, Y)$, and $X' \to U \to Y'$ is Markov.

Define the two-letter form

$$\begin{aligned}&T((X_a, X_b), (Y_a, Y_b); (X_a^0, Y_a^0), (X_b^0, Y_b^0))\\ &:= \min_{\substack{p_{U|X_a,X_b,Y_a,Y_b}:\\ p_{X_a,X_b,Y_a,Y_b|U} = p_{X_a,X_b|U}p_{Y_a,Y_b|U}}} H(X_a + Y_a, X_b + Y_b|U) - \frac{1}{2}H(X_a, X_b|U) - \frac{1}{2}H(Y_a, Y_b|U)\\ &\quad + \eta\left(H(X_a + X_a^0, X_b + X_b^0|U) - \frac{1}{2}H(X_a, X_b|U) - \frac{1}{2}H(X_a^0, X_b^0|U)\right)\\ &\quad + \eta\left(H(Y_a + Y_a^0, Y_b + Y_b^0|U) - \frac{1}{2}H(Y_a, Y_b|U) - \frac{1}{2}H(Y_a^0, Y_b^0|U)\right),\end{aligned}$$

where the tuple $(U, (X_a, X_b), (Y_a, Y_b)), X_a^0, X_b^0, Y_a^0$, and Y_b^0 are mutually independent.

Lemma 3. *For any $\eta \geq 0$, following super-additivity inequality holds:*

$$T((X_a, X_b), (Y_a, Y_b); (X_a^0, Y_a^0), (X_b^0, Y_b^0)) \geq T(X_a, Y_a; X_a^0, Y_a^0) + T(X_b, Y_b; X_b^0, Y_b^0)$$

Proof. Observe that the following holds:

$$\begin{aligned}&H(X_a + Y_a, X_b + Y_b|U) - \frac{1}{2}H(X_a, X_b|U) - \frac{1}{2}H(Y_a, Y_b|U)\\ &= H(X_a + Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b + Y_b|U, X_a - Y_a)\\ &\quad - \frac{1}{2}H(X_b|U, X_a) - \frac{1}{2}H(Y_b|U, Y_a)\\ &\overset{(a)}{=} H(X_a + Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b + Y_b|U, X_a + Y_a)\\ &\quad - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) - \frac{1}{2}H(Y_b|U, X_a, Y_a, X_a^0, Y_a^0)\\ &\geq H(X_a + Y_a|U) - \frac{1}{2}H(X_a|U) - \frac{1}{2}H(Y_a|U) + H(X_b + Y_b|U, X_a, Y_a, X_a^0, Y_a^0)\\ &\quad - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) - \frac{1}{2}H(Y_b|U, X_a, Y_a, X_a^0, Y_a^0).\end{aligned}$$

Here (a) follows from the independence and the Markov structure of the random variables.

In an identical fashion, we can also show that

$$\begin{aligned}&H(X_a + X_a^0, X_b + X_b^0|U) - \frac{1}{2}H(X_a, X_b|U) - \frac{1}{2}H(X_a^0, X_b^0|U)\\ &\geq H(X_a + X_a^0|U) - \frac{1}{2}H(X_a|U) - H(X_a^0|U) + H(X_b + X_b^0|U, X_a, Y_a, X_a^0, Y_a^0)\\ &\quad - \frac{1}{2}H(X_b|U, X_a, Y_a, X_a^0, Y_a^0) - \frac{1}{2}H(X_b^0|U, X_a, Y_a, X_a^0, Y_a^0),\end{aligned}$$

and

$$\begin{aligned}
&H(Y_a+Y_a^0, Y_b-Y_b^0|U) - \frac{1}{2}H(Y_a,Y_b|U) - \frac{1}{2}H(Y_a^0,Y_b^0|U)\\
&\geq H(Y_a+Y_a^0|U) - \frac{1}{2}H(Y_a|U) - H(Y_a^0|U) + H(Y_b+Y_b^0|U,X_a,Y_a,X_a^0,Y_a^0)\\
&\quad - \frac{1}{2}H(Y_b|U,X_a,Y_a,X_a^0,Y_a^0) - \frac{1}{2}H(Y_b^0|U,X_a,Y_a,X_a^0,Y_a^0).
\end{aligned}$$

Denote $U_a = U$, and observe that $p_{X_aY_a|U_a} = p_{X_a|U_a}p_{Y_a|U_a}$ and $(U_a, X_a, Y_a), X_a^0$, and Y_a^0 are mutually independent. Denote $U_b = (U, X_a, Y_a, X_a^0, Y_a^0)$, and observe that $p_{X_bY_b|U_b} = p_{X_b|U_b}p_{Y_b|U_b}$ and $(U_b, X_b, Y_b), X_a^0$, and Y_a^0 are mutually independent. Putting the above inequalities together, the requisite super-additivity follows. □

However, we cannot do the transformation $(X_a + X_a^0, X_b + X_b^0) \mapsto (X_a + X_a^0 + X_b + X_b^0, X_b + X_b^0)$ as this would replace X_a^0 by $X_a^0 + X_b^0$. This is not permitted as X_a^0 is a fixed distribution. Instead, one can place X_a, X_b, Y_a, Y_b at the minimizer by alternate linear forms and use the minimality to force an independence of some linear forms.

IV. A result on uniform distributions revisited

We believe that it will be illustrative to revisit the arguments in [GGMT23] in light of super-additivity. For the purpose of illustration of the ideas, we will try to keep our estimates rather elementary (the ideas are still borrowed, in many cases verbatim, from [GGMT23]). We will establish the following (weaker) result.

Theorem 7. *Let X^0, Y^0 be any pair of independent random variables with support contained inside $\mathbb{G}$, a finite Abelian group with characteristic 2. Let*

$$\begin{aligned}
\tau(X,Y) := &\left(H(X+Y) - \frac{1}{2}H(X) - \frac{1}{2}H(Y)\right) + \eta\left(H(X+X^0) - \frac{1}{2}H(X) - \frac{1}{2}H(X^0)\right)\\
&+ \eta\left(H(Y+Y^0) - \frac{1}{2}H(Y) - \frac{1}{2}H(Y^0)\right),
\end{aligned}$$

where X, Y, X^0, Y^0 are mutually independent. Here X, Y also take values in $\mathbb{G}$. Then, all minimizers of $\tau(X,Y)$ must be uniform distributions on a coset of a subgroup of $\mathbb{G}$, when $\eta \leq \eta_0$, where $\eta_0 = \frac{\sqrt{1452}-36}{26}$.

Proof of Theorem 7

We will divide the proof into some components. Some of the required inequalities will be established in the Appendix.

A. Superadditivity and rotation

Suppose (X^*, Y^*) is a minimizer of $\tau(X, Y; X^0, Y^0)$. Without loss of generality, we may assume X^* and Y^* are independent. Let (X_A, Y_A) and (X_B, Y_B) are independent copies of (X^*, Y^*). The minimality of (X^*, Y^*) implies that

$$\begin{aligned}
&\tau(X_A+U_A, Y_A+V_A; X_A^0, Y_A^0|W_A) + \tau(X_B+U_B, Y_B+V_B; X_B^0, Y_B^0|W_B)\\
&\quad\geq \tau(X_A, Y_A; X_A^0, Y_A^0) + \tau(X_B, Y_B; X_B^0, Y_B^0)
\end{aligned} \tag{1}$$

for any valid choice that $X_A + U_A \to W_A \to Y_A + V_A$ and $X_B + U_B \to W_B \to Y_B + V_B$. Here, $(U_A, V_A, W_A, U_B, V_B, W_B, X_A, Y_A, X_B, Y_B)$ is assumed to be independent of $(X_A^0, Y_A^0, X_B^0, Y_B^0)$.

Set $U_A = X_B, V_A = Y_B, W_B = (X_A + X_B, Y_A + Y_B), W_A = U_B = V_B = \emptyset$. Then (1) reduces to

$$\begin{aligned}
&I(X_A+X_B; X_B+Y_B|X_A+Y_A+X_B+Y_B)\\
&\leq \eta I(X_B; X_A+X_B+X_A^0) + \eta I(Y_B; Y_A+Y_B+Y_A^0)\\
&\qquad - \eta I(X_A+X_B; X_B+X_B^0) - \eta I(Y_A+Y_B; Y_B+Y_B^0)\\
&\leq \eta I(X_B; X_A+X_B+X_A^0) + \eta I(Y_B; Y_A+Y_B+Y_A^0)\\
&\leq \eta I(X_B; X_A+X_B) + \eta I(Y_B; Y_A+Y_B).
\end{aligned} \tag{2}$$

The last inequality is due to $(X_A^0, Y_A^0) \perp (X_A, X_B, Y_A, Y_B)$.

Similarly, by setting $U_A = Y_B, V_A = X_B, W_B = (X_A + Y_B, Y_A + X_B), W_A = U_B = V_B = \emptyset$, (1) yields

$$I(X_A+Y_B; X_B+Y_B|X_A+Y_A+X_B+Y_B) \leq \eta I(X_B; Y_A+X_B) + \eta I(Y_B; X_A+Y_B). \tag{3}$$

Finally, by setting, $U_A = X_A, V_A = Y_B, W_B = (X_A + X_B, Y_A + Y_B), W_A = U_B = V_B = \emptyset$, (1) yields

$$I(X_A+X_B; X_A+Y_B|X_A+Y_A+X_B+Y_B) \leq \eta I(X_A; X_A+X_B) + \eta I(Y_B; Y_A+Y_B). \tag{4}$$

Remark 9. We have employed three different linear transformations on the superadditive function and obtained three constraints (equations (2),(3),(4)) that has to be satisfied by the minimizer. Following the approach in the earlier sections, we need to use these inequalities to deduce some independence of linear forms, which would imply that minimizers need to be uniform. The choice of the identifications (three inequalities) above is directly motivated from [[GGMT23], Equations 3.1–3.4]. It may be possible that one could use other linear transformations and obtain additional constraints that implies the independence for a lower η but this is left for future work.

Lemma 4 ([GGMT23], Equation 5.9)**.** *Let* $S = (X_A + X_B) + (Y_A + Y_B)$.

$$H(S) - \frac{1}{2}H(X) - \frac{1}{2}H(Y) \leq (2+\eta)d(X,Y).$$

Proof. By optimality of (X, Y), we have $\tau(X_A, Y_A; X^0, Y^0|X_A + Y_B, Y_A + X_B) \geq \tau(X, Y; X^0, Y^0)$, which is equivalent to

$$\begin{aligned} d(X_A, Y_A|X_A + Y_B, Y_A + X_B) \geq d(X,Y) &- \eta(d(X^0, X_A|X_A + Y_B) - d(X^0, X_A)) \\ &- \eta(d(Y^0, Y_A|Y_A + X_B) - d(Y^0, Y_A)). \end{aligned}$$

This implies

$$\begin{aligned} &d(X_A + Y_B, Y_A + X_B) \\ &\overset{(8)}{=} 2d(X,Y) - d(X_A, Y_A|X_A + Y_B, Y_A + X_B) - I(X_A + Y_A; Y_A + X_B|X_A + X_B + Y_A + Y_B) \\ &\leq d(X,Y) + \eta(d(X^0, X_A|X_A + Y_B) - d(X^0, X_A)) + \eta(d(Y^0, Y_A|Y_A + X_B) - d(Y^0, Y_A)) \\ &\quad - I(X_A + Y_A; Y_A + X_B|X_A + X_B + Y_A + Y_B) \\ &\overset{(a)}{\leq} d(X,Y) + \eta\left(\frac{1}{2}H(X_A + Y_B) - \frac{1}{2}H(Y_B) + \frac{1}{2}H(Y_A + X_B) - \frac{1}{2}H(X_B)\right) \\ &\quad - I(X_A + Y_A; Y_A + X_B|X_A + X_B + Y_A + Y_B) \\ &= (1+\eta)d(X,Y) - I(X_A + Y_A; Y_A + X_B|X_A + X_B + Y_A + Y_B), \end{aligned}$$

where (a) follows from Family 3 of Lemma 9. Therefore,

$$d(X_A + Y_B; Y_A + X_B) \leq (1+\eta)d(X,Y).$$

This implies that

$$\begin{aligned} &H(S) - \frac{1}{2}H(X_A) - \frac{1}{2}H(Y_A) \\ &= d(X_A + Y_B; Y_A + X_B) + \frac{1}{2}I(Y_B; X_A + Y_B) + \frac{1}{2}I(X_B; Y_A + X_B) \\ &= d(X_A + Y_B; Y_A + X_B) + d(X,Y) \\ &\leq (2+\eta)d(X,Y). \end{aligned}$$

□

Lemma 5 (Inspired by [GGMT23], Equation 7.2)**.** *Let* $S = (X_A + X_B) + (Y_A + Y_B)$. *Let* $T_1 = X_A + X_B$, $T_2 = X_B + Y_B$, $T_3 = X_A + Y_B$.

$$\begin{aligned} &I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S) \\ &\qquad \leq 2\eta\Big(d(X^*, X^*) + d(Y^*, Y^*) + d(X^*, Y^*)\Big) \leq 10\eta d(X^*, Y^*). \end{aligned}$$

Proof. From the super-additivity estimation, i.e. (2),(3),(4), we have

$$\begin{aligned} &I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S) \\ &\leq \eta I(X_B; X_A + X_B) + \eta I(Y_B; Y_A + Y_B) + \eta I(X_B; Y_A + X_B) \\ &\quad + \eta I(Y_B; X_A + Y_B) + \eta I(X_A; X_A + X_B) + \eta I(Y_B; Y_A + Y_B) \\ &= \eta(2H(X_A + X_B) + 2H(Y_A + Y_B) + H(X_A + X_B) + H(X_B + Y_A) - 3H(X^*) - 3H(Y^*)) \\ &= 2\eta(d(X^*, X^*) + d(Y^*, Y^*) + d(X^*, Y^*)) \\ &\leq 10\eta d(X^*, Y^*) \end{aligned}$$

The last part of the inequality follows by the triangle inequality, Lemma 2, of the independent entropic Ruzsa distance, i.e. $d(X, X) \leq d(X, Y) + d(Y, X) = 2d(X, Y)$. □

B. Combining the estimates

We have, if (X, Y) is the minimizer for $\tau(X, Y; X_0, Y_0)$, then

$$\begin{aligned}
&\tau(X, Y; X^0, Y^0) \\
&\leq \frac{1}{6}\Bigg(\tau(T_1, T_2; X^0, Y^0|T_3, S) + \tau(T_2, T_3; X^0, Y^0|T_1, S) + \tau(T_3, T_1; X^0, Y^0|T_2, S) \\
&\quad + \tau(T_2, T_1; X^0, Y^0|T_3, S) + \tau(T_3, T_2; X^0, Y^0|T_1, S) + \tau(T_1, T_3; X^0, Y^0|T_2, S)\Bigg) \\
&= \frac{1}{3}\Bigg(d(T_1, T_2|T_3, S) + d(T_2, T_3|T_1, S) + d(T_3, T_1|T_2, S)\Bigg) \\
&\quad + \frac{\eta}{6}\sum_{i=1}^{3}\sum_{j\neq i}\Bigg(d(X^0, T_i|T_j, S) + d(Y^0, T_i|T_j, S)\Bigg) \\
&\overset{(a)}{\leq} I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S) + \frac{\eta}{3}\sum_{i=1}^{3}\Bigg(d(X^0, T_i|S) + d(Y^0, T_i|S)\Bigg) \\
&\quad + \frac{\eta}{3}\Bigg(I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S)\Bigg) \\
&\overset{(b)}{\leq} \left(1 + \frac{\eta}{3}\right)\Bigg(I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S)\Bigg) \\
&\quad + \eta\Bigg(d(X, X^0) + d(Y, Y^0) + H(S) - \frac{1}{2}H(X) - \frac{1}{2}H(Y)\Bigg),
\end{aligned}$$

where (a) follows by Corollary 1 and Lemma 7, (b) follows by Families 1 and 2 of Lemma 9.

By the definition of the PFR functional, we have

$$\begin{aligned}
d(X^*, Y^*) &\leq \left(1 + \frac{\eta}{3}\right)\Bigg(I(T_1; T_2|S) + I(T_2; T_3|S) + I(T_1; T_3|S)\Bigg) \\
&\quad + \eta\Bigg(H(S) - \frac{1}{2}H(X) - \frac{1}{2}H(Y)\Bigg) \\
&\leq \left(1 + \frac{\eta}{3}\right)10\eta d(X^*, Y^*) + \eta(2 + \eta)d(X^*, Y^*),
\end{aligned}$$

where the last inequality is a consequence of Lemma 4 and Lemma 5. Therefore, if $1 > \left(1 + \frac{\eta}{3}\right)10\eta + \eta(2 + \eta)$, for some $\eta > 0$, then $d(X^*, Y^*) = 0$. Note that $\eta_0 = \frac{\sqrt{1452}-36}{26}$, is the positive root of $1 = \left(1 + \frac{\eta}{3}\right)10\eta + \eta(2 + \eta)$. Therefore, for $\eta < \eta_0$, $d(X^*, Y^*) = 0$, or in other words, $0 = H(X^* + Y^*) - \frac{1}{2}H(X^*) - \frac{1}{2}H(Y^*) = \frac{1}{2}I(X^*; X^* + Y^*) + \frac{1}{2}I(Y^*; X^* + Y^*)$. This implies that X^* is independent of $X^* + Y^*$ and Y^* is independent of $X^* + Y^*$.

C. Completing the proof of Theorem 7

From Lemma 1 and that X^* is independent of $X^* + Y^*$, we have $|\text{support}(X^*)| \geq k|\text{support}(Y^*)|$ for some $k \in \mathbb{N}$, and from Y^* is independent of $X^* + Y^*$, we have $|\text{support}(Y^*)| \geq k|\text{support}(X^*)|$, $k \in \mathbb{N}$. This implies that $|\text{support}(X^*)| = |\text{support}(Y^*)|$. Further, from Lemma 1, we can also conclude that $|\mathbb{D}| = |\text{support}(Y^*)|$, where $\mathbb{D}$ denote the subgroup generated by pairwise differences of the elements of $\text{support}(Y^*)$. This implies that Y^* is supported on a coset of $\mathbb{D}$. Further, from Lemma 1, and $|\text{support}(X^*)| = |\mathbb{D}|$, we can also infer that X^* is uniform on a coset of $\mathbb{D}$. Reversing the roles of X^* and Y^*, we can also infer the Y^* is uniform over its support. Therefore, X^* and Y^* are uniformly distributed on cosets of the same subgroup.

V. Future and Related Work

There are extensions of the EPI that are known in the literature. One of the extensions is as follows,

Theorem 8. *Suppose $X_1, \ldots, X_n$ are independent continuous random variables with finite variables, and let $(\alpha_1, \ldots, \alpha_n)$ be non-negative constants such that $\sum_{i=1}^{n}\alpha_i = 1$. We have*

$$nh\left(\sum_{i=1}^{n}\sqrt{\alpha_i}X_i\right) \geq \sum_{j=1}^{n}(1 - \alpha_j)h\left(\sum_{i\neq j}\sqrt{\frac{\alpha_i}{1 - \alpha_j}}X_i\right).$$

This formulation is established by Artstein, Ball, Barthe, and Naor [ABBN04]. Johnson and Yu proposed a discrete analog of this result using the Rényi thinning of discrete random variables [JY10].

An extension to torsion-free groups is certainly interesting. Along these lines, Tao proposed a conjecture [Tao10] on the discrete analog of EPI as below

Conjecture 1. *Suppose $X_1, \ldots, X_{n+1}$ are identically distributed and independent random variables on some torsion-free group $\mathbb{T}$. Then, for any $\epsilon > 0$, as long as $H(X)$ is sufficiently large (depending on n, ϵ), we have*

$$H(X_1 + \cdots + X_{n+1}) \geq H(X_1 + \cdots + X_n) + \frac{1}{2} \log \frac{n+1}{n} - \epsilon.$$

There is a recent proof of this conjecture, under the assumption that the distribution of X is log-concave, by Gavalakis [Gav23].

In a problem related to the capacity region of a Gaussian noise channel with Z-interference, there is an information functional, similar to the one studied in the second half. It is conjectured that Gaussian distributions are optimal (similar to uniforms) under some parameter regimes (just like in this setting where uniforms are optimal only when η is small enough). Perhaps the ideas employed here can be adapted to that setting, and thereby solving a long-standing open problem in multiuser information theory.

REFERENCES

[ABBN04] Shiri Artstein, Keith M. Ball, Franck Barthe, and Assaf Naor, *Solution of Shannon's problem on the monotonicity of entropy*, Journal of the American Mathematical Society **17** (2004), no. 4, 975–982 (English (US)).

[AJN22] Venkat Anantharam, Varun Jog, and Chandra Nair, *Unifying the Brascamp-Lieb inequality and the entropy power inequality*, IEEE Transactions on Information Theory **68** (2022), no. 12, 7665–7684.

[BCCT08] J. Bennett, A. Carbery, M. Christ, and T. Tao, *The Brascamp-Lieb inequalities: Finiteness, structure and extremals*, Geometric and Functional Analysis **17** (2008), no. 5, 1343–1415.

[Ber73] P F Bergmans, *Coding theorem for broadcast channels with degraded components*, IEEE Trans. Info. Theory **IT-15** (March, 1973), 197–207.

[BL76] H. J. Brascamp and E. H. Lieb, *Best constants in Young's inequality, its converse, and its generalization to more than three functions*, Advances in Mathematics **20** (1976), no. 2, 151–173.

[Dar53] G. Darmois, *Analyse générale des liaisons stochastiques: etude particulière de l'analyse factorielle linéaire*, Revue de l'Institut International de Statistique / Review of the International Statistical Institute **21** (1953), no. 1/2, pp. 2–8 (English).

[Fel99] Gennadiy Feldman, *More on the Skitovich-Darmois theorem for finite abelian groups*, Theory of Probability and Its Applications **45** (1999).

[Gav23] Lampros Gavalakis, *Discrete generalised entropy power inequalities for log-concave random variables*, 2023 IEEE International Symposium on Information Theory (ISIT), 2023, pp. 42–47.

[GGMT23] WT Gowers, Ben Green, Freddie Manners, and Terence Tao, *On a conjecture of Marton*, arXiv preprint arXiv:2311.05762 (2023).

[GN14] Y. Geng and C. Nair, *The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages*, IEEE Transactions on Information Theory **60** (2014), no. 4, 2087–2104.

[HV03] Peter Harremoës and CHRISTOPHE VIGNAT, *An entropy power inequality for the binomial family*, JIPAM. Journal of Inequalities in Pure & Applied Mathematics [electronic only] **4** (2003).

[JA14] Varun Jog and Venkat Anantharam, *The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n*, IEEE Transactions on Information Theory **60** (2014), no. 7, 3773–3786.

[JY10] Oliver Johnson and Yaming Yu, *Monotonicity, thinning, and discrete versions of the entropy power inequality*, IEEE Transactions on Information Theory **56** (2010), no. 11, 5387–5395.

[KLN23] Chin Wa Ken Lau and Chandra Nair, *Information inequalities via ideas from additive combinatorics*, 2023 IEEE International Symposium on Information Theory (ISIT), 2023, pp. 2452–2457.

[Lie78] Elliott H. Lieb, *Proof of an entropy conjecture of Wehrl*, Comm. Math. Phys. **62** (1978), no. 1, 35–41.

[LN24] Chin Wa Lau and Chandra Nair, *An entropic inequality in finite abelian groups analogous to the unified Brascamp-Lieb and entropy power inequality*, 2024 IEEE International Symposium on Information Theory (ISIT), IEEE, 2024, pp. 3588–3593.

[SDM11] Naresh Sharma, Smarajit Das, and Siddharth Muthukrishnan, *Entropy power inequality for a family of discrete random variables*, 2011 IEEE International Symposium on Information Theory Proceedings, 2011, pp. 1945–1949.

[Sha48] C E Shannon, *A mathermatical theory of communication*, Bell System Technical Journal **27** (July and October, 1948), 379–423 and 623–656.

[Ski53] Viktor P Skitovitch, *On a property of the normal distribution*, DAN SSSR **89** (1953), 217–219.

[Sta59] A.J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), no. 2, 101–112.

[SW90] S. Shamai and A.D. Wyner, *A binary analog to the entropy-power inequality*, IEEE Transactions on Information Theory **36** (1990), no. 6, 1428–1430.

[Tao10] Terence Tao, *Sumset and inverse sumset theory for Shannon entropy*, Combinatorics, Probability and Computing **19** (2010), no. 4, 603–639.

[WSS06] H. Weingarten, Y. Steinberg, and S. S. Shamai, *The capacity region of the Gaussian multiple-input multiple-output broadcast channel*, IEEE Transactions on Information Theory **52** (2006), no. 9, 3936–3964.

[WZ73] A. Wyner and J. Ziv, *A theorem on the entropy of certain binary sequences and applications: Part I*, IEEE Trans. Inform. Theory **IT-19** (1973), no. 6, 769–772.

APPENDIX

Lemma 6 ([KLN23])**.** *Let $(X_i)_{i=1}^n$ be a sequence of finite-valued random variables (defined on some common probability space) and $(f_i, g_i)_{i=1}^{n-1}$ be a sequence of functions that take a finite set of values in some space $\mathcal{S}$ such that: $f_i(X_i) = g_i(X_{i+1})(=: U_i)$ and the following Markov chain holds,*

$$X_1 \to U_1 \to X_2 \to U_2 \to \cdots \to X_{n-1} \to U_{n-1} \to X_n.$$

Then,

$$H(X_1,\dots,X_n)+\sum_{i=1}^{n-1}H(U_i)=\sum_{i=1}^{n}H(X_i).$$

Proof. Note that $H(X_1,\dots,X_n)=H(X_1,\dots,X_n,U_1,\dots,U_{n-1})$ since U_i is determined by X_i (and also by X_{i+1}). Now observe that

$$\begin{aligned}
&H(X_1,\dots,X_n)+\sum_{i=1}^{n-1}H(U_i)\\
&\overset{(a)}{=} H(X_1,\dots,X_n,U_1,\dots,U_{n-1})+\sum_{i=1}^{n-1}H(U_i)\\
&\overset{(b)}{=} H(X_1,U_1)+H(X_2,U_2|X_1,U_1)+\cdots+H(X_{n-1},U_{n-1}|X_1^{n-2},U_1^{n-2})\\
&\quad+H(X_n|X_1^{n-1},U_1^{n-1})+\sum_{i=1}^{n-1}H(U_i)\\
&\overset{(c)}{=} H(X_1,U_1)+H(X_2,U_2|U_1)+\cdots+H(X_{n-1},U_{n-1}|U_{n-2})+H(X_n|U_{n-1})+\sum_{i=1}^{n-1}H(U_i)\\
&\overset{(d)}{=} H(X_1,U_1)+H(X_2,U_2,U_1)+\cdots+H(X_{n-1},U_{n-1},U_{n-2})+H(X_n,U_{n-1})\\
&\overset{(e)}{=} \sum_{i=1}^{n}H(X_i).
\end{aligned}$$

In the above (a) and (e) follow using the assumption that U_i is determined by any of X_i or X_{i+1}. The equalities (b) and (d) are a consequence of the chain rule for entropy and the equality (c) is a consequence of the Markov chain assumption. □

A. Preliminary inequalities

In this section, we present some preliminary inequalities that were (essentially) established in [GGMT23].

Lemma 7 (Adapted from [GGMT23], Lemma 5.2)**.** *Suppose X is independent of Y, Z, we have*

$$d(X,Y|Z)\le d(X,Y)+\frac{1}{2}I(Y;Z).$$

Proof. We have

$$\begin{aligned}
d(X,Y|Z)&=H(X+Y|Z)-\frac{1}{2}H(X)-\frac{1}{2}H(Y|Z)\\
&=d(X,Y)+\frac{1}{2}I(Y;Z)-I(X+Y;Z).
\end{aligned}$$

□

Lemma 8 ([GGMT23], Lemma A.2)**.** *Let A,B,S be jointly distributed on $\mathbb{G}$.*

$$d(A,B|Z,S)\le 3I(A;B|S)+2H(A-B|S)-H(A|S)-H(B|S).$$

Proof. Let $Z=A-B$. Given S, construct two "copies" of (A,B), labeled as (A_1,B_1) and (A_2,B_2) such that $A_1-B_1=Z=A_2-B_2$ and their joint law satisfies $p_S p_{Z|S} p_{A_1,B_1|Z,S} p_{A_2,B_2|Z,S}$.

We have (from sub-modularity)

$$\begin{aligned}
H(A_1+B_2,A_1,B_1|S)+H(A_1+B_2|S)&\le H(A_1+B_2,A_1|S)+H(A_1+B_2,B_1|S)\\
&=H(A_1,B_2|S)+H(A_2+B_1,B_1|S)\\
&=H(A_1,B_2|S)+H(A_2,B_1|S) \qquad (5)
\end{aligned}$$

Copy lemma, Lemma 6, yields

$$H(A_1,A_2,B_1,B_2|S)+H(A_1-B_1|S)=H(A_1,B_1|S)+H(A_2,B_2|S)$$

However we also have that (A_1+B_2,A_1,B_1) determines and is determined by A_1,B_1,A_2,B_2. Therefore

$$\begin{aligned}
H(A_1,B_1|S)+H(A_2,B_2|S)&=H(A_1,B_1,A_2,B_2|S)+H(A_1-B_1|S)\\
&=H(A_1+B_2,A_1,B_1|S)+H(A_1-B_1|S)
\end{aligned}$$

$$\overset{(5)}{\leq} H(A_1, B_2|S) + H(A_2, B_1|S) - H(A_1 + B_2|S) + H(A_1 - B_1|S)$$

Rearranging yields,

$$H(A_1 + B_2|S) \leq H(A_1, B_2|S) + H(A_2, B_1|S) - H(A_1, B_1|S) - H(A_2, B_2|S) + H(A_1 - B_1|S). \tag{6}$$

We also have $H(A|A-B,S) = H(B|A-B,S) = H(A,B|S) - H(A-B|S)$. Therefore,

$$\begin{aligned}
&H(A_1 + B_2|S) - \frac{1}{2}H(A_1|A_1 - B_1, S) - \frac{1}{2}H(B_2|A_2 - B_2, S) \\
&= H(A_1 + B_2|S) - \frac{1}{2}H(A_1, B_1|S) - \frac{1}{2}H(A_2, B_2|S) + H(A - B|S) \\
&\overset{(6)}{\leq} H(A_1, B_2|S) + H(A_2, B_1|S) - \frac{3}{2}H(A_1, B_1|S) - \frac{3}{2}H(A_2, B_2|S) + 2H(A - B|S) \\
&\leq 3I(A; B|S) + 2H(A - B|S) - H(A|S) - H(B|S).
\end{aligned} \tag{7}$$

From definition

$$\begin{aligned}
d(A, B|Z, S) &= H(A_1 + B_2|Z, S) - \frac{1}{2}H(A_1|Z, S) - \frac{1}{2}H(B_2|Z, S) \\
&\leq H(A_1 + B_2|S) - \frac{1}{2}H(A_1|Z, S) - \frac{1}{2}H(B_2|Z, S).
\end{aligned}$$

Equation (7) completes the proof. □

Corollary 1 (From the arguments in [GGMT23], Lemma 7.2)**.** *Let (S, T_1, T_2, T_3) be jointly distributed on a group of characteristic two such that $T_1 + T_2 + T_3 = 0$. Then, we have*

$$\begin{aligned}
&d(T_2, T_3|T_1, S) + d(T_3, T_1|T_2, S) + d(T_1, T_2|T_3, S) \\
&\quad \leq 3I(T_1; T_2|S) + 3I(T_2; T_3|S) + 3I(T_1; T_3|S)
\end{aligned}$$

Proof. Under the characteristic two assumption, Lemma 8 yields

$$\begin{aligned}
d(T_2, T_3|T_1, S) &\leq 3I(T_1; T_2|S) + 2H(T_2 - T_3|S) - H(T_2|S) - H(T_3|S) \\
&= 3I(T_1; T_2|S) + 2H(T_2 + T_3|S) - H(T_2|S) - H(T_3|S) \\
&= 3I(T_1; T_2|S) + 2H(T_1|S) - H(T_2|S) - H(T_3|S).
\end{aligned}$$

The corollary follows by adding the cyclic shifts of this inequality. □

Lemma 9 (Adapted from [GGMT23], Lemma 7.1 and Section 7)**.** *Let X and Y be two independent random variables defined on a field of characteristic two. Let (X_A, Y_A) and (X_B, Y_B) be independent copies of (X, Y). Let $S = (X_A + X_B) + (Y_A + Y_B)$. Let $T_1 = (X_A + X_B)$, $T_2 = (X_B + Y_B)$, $T_3 = X_A + Y_B$. Let (X^0, Y^0) be independent of the other random variables.*

1) Family 1: The following inequalities hold

$$\begin{aligned}
d(X^0, T_2|S) - d(X^0, X) &\leq \frac{1}{2}H(S) - \frac{1}{2}H(X), \\
d(X^0, T_3|S) - d(X^0, X) &\leq \frac{1}{2}H(S) - \frac{1}{2}H(X), \\
d(Y^0, T_2|S) - d(Y^0, Y) &\leq \frac{1}{2}H(S) - \frac{1}{2}H(Y), \\
d(Y^0, T_3|S) - d(Y^0, Y) &\leq \frac{1}{2}H(S) - \frac{1}{2}H(Y),
\end{aligned}$$

2) Family 2: The following inequalities hold

$$\begin{aligned}
d(X^0, T_1|S) - d(X^0, X) &\leq \frac{1}{2}H(S) + \frac{1}{2}H(X_A + X_B) - \frac{1}{2}H(Y_A + Y_B) - \frac{1}{2}H(X) \\
d(Y^0, T_1|S) - d(Y^0, Y) &\leq \frac{1}{2}H(S) + \frac{1}{2}H(Y_A + Y_B) - \frac{1}{2}H(X_A + X_B) - \frac{1}{2}H(Y).
\end{aligned}$$

3) Family 3: The following inequalities hold

$$\begin{aligned}
d(X^0, X_A|X_A + Y_B) - d(X^0, X_A) &\leq \frac{1}{2}H(X_A + Y_B) - \frac{1}{2}H(Y_B), \\
d(Y^0, Y_A|Y_A + X_B) - d(Y^0, Y_A) &\leq \frac{1}{2}H(Y_A + X_B) - \frac{1}{2}H(X_B).
\end{aligned}$$

4) The following equality holds:

$$\begin{aligned} &d(X_A+Y_B,Y_A+X_B)+d(X_A,Y_A|X_A+Y_B,Y_A+X_B)\\ &\quad+I(X_A+Y_A;Y_A+X_B|X_A+X_B+Y_A+Y_B)\\ &\quad=2d(X,Y). \end{aligned} \tag{8}$$

Proof. Family 1: The proofs of the inequalities in the first family are similar. We provide the details of the first. It suffices to show that

$$H(X_B+Y_B+X^0|S)-\frac{1}{2}H(X_B+Y_B|S)-H(X+X^0)+\frac{1}{2}H(X)\leq\frac{1}{2}H(S)-\frac{1}{2}H(X).$$

This is equivalent to showing that

$$H(X_B+Y_B+X^0|S)-\frac{1}{2}H(X_B+Y_B,S)\leq I(X^0;X+X^0).$$

Observe that

$$H(X_B+Y_B,S)=H(X_A+Y_A)+H(X_B+Y_B)=2H(X_B+Y_B).$$

Therefore, it suffices to show that

$$H(X_B+Y_B+X^0|S)-H(X_B+Y_B)\leq I(X^0;X+X^0)$$

Note that

$$\begin{aligned} H(X_B+Y_B+X^0|S)-H(X_B+Y_B)&\leq H(X_B+Y_B+X_0)-H(X_B+Y_B)\\ &=I(X_0;X_B+Y_B+X_0)\leq I(X_0;X+X_0). \end{aligned}$$

The last inequality is a consequence of the data-processing inequality as Y_B is independent of (X_0,X_B). This establishes the first inequality.

Family 2: The proofs of the inequalities in the second family are similar. We only prove the first one. We wish to show that

$$\begin{aligned} &H(T_1+X^0|S)-\frac{1}{2}H(T_1|S)-H(X+X^0)+\frac{1}{2}H(X)\\ &\quad\leq\frac{1}{2}H(S)+\frac{1}{2}H(X_A+X_B)-\frac{1}{2}H(Y_A+Y_B)-\frac{1}{2}H(X), \end{aligned}$$

or equivalently

$$H(T_1+X^0|S)\leq\frac{1}{2}H(T_1,S)+\frac{1}{2}H(X_A+X_B)-\frac{1}{2}H(Y_A+Y_B)+H(X+X^0)-H(X),$$

Since $H(T_1,S)+H(S)=H(X_A+X_B,Y_A+Y_B)=H(X_A+X_B)+H(Y_A+Y_B)$ (the random variables are independent),we wish to show that

$$H(T_1+X^0|S)\leq H(X_A+X_B)+H(X+X^0)-H(X).$$

Therefore, it suffices to prove the stronger inequality that

$$H(T_1+X^0)\leq H(X_A+X_B)+H(X+X^0)-H(X),$$

or equivalently

$$I(X^0;X_A+X_B+X^0)\leq I(X^0;X_A+X^0)=I(X^0;X+X^0).$$

This inequality is a consequence of the data-processing inequality as X_B is independent of (X_0,X_A). This establishes the desired inequality.

Family 3: The proofs of the inequalities in the third family are similar. We only prove the first one. We wish to show that

$$H(X^0+X_A|X_A+Y_B)-\frac{1}{2}H(X_A|X_A+Y_B)-H(X^0+X_A)+\frac{1}{2}H(X_A)\leq\frac{1}{2}H(X_A+Y_B)-\frac{1}{2}H(Y_B).$$

This is equivalent to showing that

$$H(X^0+X_A|X_A+Y_B)-H(X^0+X_A)\leq\frac{1}{2}H(X_A|X_A+Y_B)+\frac{1}{2}H(X_A+Y_B)-\frac{1}{2}H(X_A)-\frac{1}{2}H(Y_B).$$

Note that $H(X_A|X_A+Y_B)+H(X_A+Y_B)=H(X_A,Y_B)=H(X_A)+H(Y_B)$. Therefore, the right-hand-side of the desired inequality is zero. On the other hand, $H(X^0+X_A|X_A+Y_B)\leq H(X^0+X_A)$ is immediate, establishing the desired inequality.

Final Equality: Observe that

$$d(X_A+Y_B,Y_A+X_B)+d(X_A,Y_A|X_A+Y_B,Y_A+X_B)$$

$$
\begin{aligned}
&\quad + I(X_A + Y_A; Y_A + X_B | X_A + X_B + Y_A + Y_B) \\
&= H(X_A + Y_B + Y_A + X_B) - \frac{1}{2}H(X_A + Y_B) - \frac{1}{2}H(Y_A + X_B) \\
&\quad + H(X_A + X_B | X_A + Y_B, Y_A + X_B) - \frac{1}{2}H(X_A | X_A + Y_B) - \frac{1}{2}H(Y_A | Y_A + X_B) \\
&\quad + I(X_A + Y_A; Y_A + X_B | X_A + X_B + Y_A + Y_B) \\
&= H(X_A + Y_B + Y_A + X_B) + H(X_A + X_B | X_A + Y_B, Y_A + X_B) \\
&\quad - \frac{1}{2}H(X_A, X_A + Y_B) - \frac{1}{2}H(Y_A, Y_A + X_B) \\
&\quad + H(X_A + Y_A | X_A + X_B + Y_A + Y_B) - H(X_A + Y_A | X_A + Y_B, Y_A + X_B) \\
&= H(X_A + Y_A, X_B + Y_B) - \frac{1}{2}H(X_A, Y_B) - \frac{1}{2}H(Y_A, X_B) = 2d(X, Y).
\end{aligned}
$$

□